

Aperiodic two-dimensional words of small abelian complexity

Svetlana Puzynina

Saint Petersburg State University
Saint Petersburg, Russia

Sobolev Institute of Mathematics
Novosibirsk, Russia

s.puzynina@gmail.com

Submitted: Mar 13, 2019; Accepted: Sep 9, 2019; Published: Oct 11, 2019

© The author.

Abstract

In this paper we prove an abelian analog of the famous Nivat's conjecture linking complexity and periodicity for two-dimensional words: We show that if a two-dimensional recurrent word contains at most two abelian factors for each pair (n, m) of integers, then it has a periodicity vector. Moreover, we show that a two-dimensional aperiodic recurrent word must have more than two abelian factors infinitely often. On the other hand, there exist aperiodic recurrent words with abelian complexity bounded by 3, as well as aperiodic words having abelian complexity 1 for some pairs (m, n) .

Mathematics Subject Classifications: 68R15, 37B50

1 Introduction

Nivat's conjecture, introduced at ICALP 1997, is a two-dimensional analog of a classical one-dimensional theorem of Morse and Hedlund. In their seminal paper *Symbolic Dynamics* (1938), besides giving formal foundations to a new mathematical field, they introduced the notion of complexity of an infinite word as a function $p(n)$ which counts, for each integer n , the number of its distinct factors (i.e., blocks of consecutive letters) of length n . The factor complexity provides a useful measure of randomness of an infinite word and more generally of the subshift it generates. They further proved that an infinite word is periodic if and only if its complexity function satisfies $p(n) \leq n$ for each n . The latter result is referred to as Morse-Hedlund theorem.

Nivat's conjecture is a generalization of Morse-Hedlund theorem to two dimensions. A two-dimensional word w (or a configuration) is an element of $\Sigma^{\mathbb{Z}^2}$, where Σ is a finite

set called an alphabet. In other words, this is a mapping w assigning a symbol from the alphabet Σ to every vertex of the lattice \mathbb{Z}^2 . A word w is called *periodic* if there exists a vector $\mathbf{v} \in \mathbb{Z}^2$ such that for every point $\mathbf{x} \in \mathbb{Z}^2$ we have $w(\mathbf{x} + \mathbf{v}) = w(\mathbf{x})$. A complexity in two dimensions is a function $p_w(m, n)$ counting for each $m, n \in \mathbb{N}$ the number of distinct rectangular $m \times n$ blocks.

Conjecture 1. (Nivat’s conjecture [20]) Let w be a two-dimensional word, n, m two numbers such that the complexity p_w satisfies $p_w(n, m) \leq nm$. Then w has a periodicity vector.

Weak forms of the conjecture for $p_w(n, m) \leq \frac{nm}{c}$ for different constants $c > 0$ were proved by Epifanio, Koskas, Mignosi [12], by Quas, Zamboni [24], by Cyr, Kra [5], and in asymptotic form by Kari, Szabados [7]; minimal complexity of aperiodic two-dimensional words was explored by Berthé and Vuillon [3]. Remarkably, the conjecture does not hold in higher dimensions [10]. About Nivat’s conjecture with an alternative notion for periodicity, see [6].

The problem to be studied concerns abelian complexity of multidimensional words and in particular an abelian analog of Nivat’s conjecture under recurrence condition. Various abelian properties of (one-dimensional) words, including abelian complexity, have been widely studied recently, see, e.g. [1, 16, 21, 22, 25]. A closely related paper studies balance in two-dimensional words [2]. It is worth noting that, contrary to the one-dimensional case, in two dimensions balance is not equivalent to bounded abelian complexity.

The paper is organized as follows. In Section 2, we provide some background and give some definitions and notation. In Section 3 we provide our main results. In Subsection 3.1 we show that, contrary to the one-dimensional case, there exist aperiodic words with abelian complexity 1 for some block sizes. In Subsection 3.2, we show that for recurrent words abelian complexity cannot be bounded by 2. After providing some observations about small abelian complexity of one-dimensional words in Subsection 3.3, we show in Subsection 3.4 that for aperiodic recurrent two-dimensional words abelian complexity at least 3 must be achieved for infinitely many block sizes.

2 Preliminaries

2.1 One-dimensional words

Let Σ be a finite non-empty set. For each infinite word $w = w_0w_1w_2 \cdots \in \Sigma^{\mathbb{N}}$, the complexity or *factor complexity* $p_w(n)$ counts the number of distinct blocks $w_iw_{i+1} \cdots w_{i+n-1} \in \Sigma^n$ of length n occurring in w . In a similar way the complexity function can be defined for a biinfinite word $w \in \Sigma^{\mathbb{Z}}$. First introduced by Hedlund and Morse in their seminal 1938 paper [18] under the name of *block growth*, the factor complexity provides a useful measure of the extent of randomness of x and more generally of the subshift it generates. A celebrated theorem of Morse and Hedlund in [18] states that every aperiodic word contains at least $n + 1$ distinct factors of each length n .

An infinite word w is called *recurrent* if each of its factors occurs in it infinitely many times. An infinite word w is called *uniformly recurrent* if for each integer n there exists an

integer N such that each of factors of w of length N contains all factors of w of length n . In other words, a word is uniformly recurrent if each its factor occurs in it with bounded gap. An infinite word w is called *eventually periodic* if there exist integers N and T such that $w_{n+T} = w_n$ for each $n \geq N$. An infinite word w is called *purely periodic* if $w_{n+T} = w_n$ for each $n \geq 0$. A word is *aperiodic* if it is not ultimately periodic. A biinfinite word w is *periodic* if $w_{n+T} = w_n$ for each integer n and aperiodic otherwise. For more on periodicity we refer to [26].

For a finite word v and a letter a , we let $|v|_a$ denote the number of occurrences of a in v . The *frequency* of a in v is $\text{Freq}_a(v) = \frac{|v|_a}{|v|}$. A (bi-)infinite word w has *uniform frequency* of a letter a if the ratio $\frac{|w_k \cdots w_{k+n-1}|_a}{n}$ has a limit $\text{Freq}_a(w)$ when $n \rightarrow \infty$, uniformly in k . An infinite word is called *balanced* if each its two factors u and v of the same length satisfy $||u|_a - |v|_a| \leq 1$ for each letter $a \in \Sigma$.

Theorem 2 (Morse and Hedlund, 1938). *Let w be a one-dimensional word. If there exists n such that $p_w(n) \leq n$, then w is eventually periodic.*

Words satisfying $p_x(n) = n + 1$ for each $n \geq 0$ are called Sturmian words, and hence are regarded as the simplest aperiodic words. Sturmian words admit various types of characterizations of geometric and combinatorial nature, e.g., they can be defined via balance, complexity, morphisms, etc. (see Chapter 2 in [15]).

In this paper we are interested in extending the result of Morse and Hedlund to the abelian setting. Two finite words u and v are *abelian equivalent* if for each letter a , the number of occurrences of a in u (denoted $|u|_a$) is equal to $|v|_a$. In other words, u and v are permutations of one another. It is straightforward that abelian equivalence is indeed an equivalence relation on the set of finite words. For each infinite word $w = w_0w_1w_2 \cdots \in \Sigma^{\mathbb{N}}$, the *abelian complexity* $a_w(n)$ counts the number of distinct abelian classes of blocks $w_iw_{i+1} \cdots w_{i+n-1} \in \Sigma^n$ of length n occurring in w . Ordering the alphabet $\Sigma = \{a_1, \dots, a_k\}$, we define Parikh vector of a finite word v over Σ as $PV(v) = (|v|_{a_1}, \dots, |v|_{a_k})$. The set of Parikh vectors of factors of length n of an infinite word w is then denoted by $PV_w(n)$ (the index w is omitted when no ambiguity arises).

Lemma 3. *Let w be a one-dimensional word. If there exists n such that $a_w(n) = 1$, then w is purely periodic.*

On the other hand, Sturmian words are aperiodic and have abelian complexity 2 for each n , and moreover this is a characterization:

Theorem 4 ([19]). *Sturmian words are aperiodic one-dimensional words with abelian complexity $a_w(n) = 2$ for each integer n .*

Therefore, Sturmian words are aperiodic words of minimal abelian complexity as well. We remark that the above results and definitions extend to biinfinite words.

2.2 Two-dimensional words

A two-dimensional word w is called *periodic* if there exists a pair of integers (m, n) such that $w(x, y) = w(x + m, y + n)$ for each pair (x, y) of integers. A two-dimensional word w is called *fully periodic* if there exist two integer vectors $(m_1, n_1), (m_2, n_2)$ linearly independent over \mathbb{Q} , such that $w(x, y) = w(x + m_i, y + n_i)$ for each pair (x, y) of integers and $i = 1, 2$. By an $m \times n$ factor (or block) of a two-dimensional word w we mean a rectangular fragment of w of the form

$$\begin{array}{ccc} w_{x,y+n-1} & \cdots & w_{x+m-1,y+n-1} \\ \vdots & & \vdots \\ w_{x,y} & \cdots & w_{x+m-1,y} \end{array}$$

for some integers x, y . We say that the block occurs at position (x, y) and we let $B_{m,n}(x, y)$ denote such a block.

Similarly to the one-dimensional case, the *abelian complexity* $a_w(m, n)$ of a two-dimensional word w is defined as the number of abelian classes of $m \times n$ blocks. A two-dimensional word is *balanced* if for each pair (m, n) of integers, each letter a and any two $m \times n$ -factors u and v of the words it holds $||u|_a - |v|_a| \leq 1$. Sometimes C -balance is defined if the equality $||u|_a - |v|_a| \leq C$ holds. We remark that although in the one-dimensional case bounded abelian complexity is equivalent to C -balance for some constant C , in two dimensions it is not true: Consider, for example, a word with alternating horizontal lines of 1's and of 0's. This word has bounded abelian complexity but is not balanced.

A two-dimensional word w is called *recurrent* if each its factor occurs in it infinitely many times. An infinite word w is called *uniformly recurrent* if for each integer n there exists an integer N such that each square $N \times N$ factor of w contains all square factors of w of size $n \times n$. For different versions of recurrence in 2D-words, we refer to [4].

We now need a few technical definitions. We call an (m, n) -lattice rooted in $(x, y) \in \mathbb{Z}^2$ the set $\{(x, y) + (mi, nj) | i, j \in \mathbb{Z}\}$. We remark that we can assume that $0 \leq x < m, 0 \leq y < n$, and that this way \mathbb{Z}^2 is split into mn many (m, n) -lattices.

A two-dimensional infinite word w has *uniform frequency of a letter* a if the ratio $\frac{|B_{m,m}(x,y)|_a}{m^2}$ has a limit $\text{Freq}_a(w)$ when $m \rightarrow \infty$, uniformly in (x, y) . We will also use the notation $PV_w(m, n)$ for the set of Parikh vectors of $m \times n$ -blocks of a word w (the index w is omitted when no ambiguity arises). The abelian complexity then satisfies $a_w(m, n) = \#PV_w(m, n)$.

3 Small abelian complexity of two-dimensional words

In this section we provide our main results. In the first subsection we consider the case of abelian complexity equal to 1 for some blocks. In particular, we study the structure of such words and show that, contrary to the one-dimensional case, there exist aperiodic words with abelian complexity 1 for some block sizes. In the second subsection we show that for recurrent two-dimensional words abelian complexity cannot be bounded by 2, but

there exist aperiodic recurrent two-dimensional words with abelian complexity bounded by 3. In the third subsection, we give some observations about small abelian complexity of one-dimensional words. Finally, in the last subsection we show that for aperiodic recurrent two-dimensional words abelian complexity at least 3 must be achieved for an infinite number of block sizes.

3.1 Abelian complexity 1 from some block size

First we describe the structure of two-dimensional words having abelian complexity 1 for some block size:

Theorem 5. *Let w be a two-dimensional word, and let $a_w(m, n) = 1$ for some integers m and n . Then in each (m, n) -lattice w is either $(0, n)$ -periodic or $(m, 0)$ -periodic, i.e., we have either $w(x, y) = w(x + m, y)$ for each point (x, y) from the lattice, or $w(x, y) = w(x, y + n)$ for each point (x, y) from the lattice.*

Proof. Consider an (m, n) -lattice and any point $(x, y) \in \mathbb{Z}^2$ from the lattice. Considering Parikh vectors of four $m \times n$ rectangles $B_{m,n}(x, y)$, $B_{m,n}(x + 1, y)$, $B_{m,n}(x, y + 1)$, $B_{m,n}(x + 1, y + 1)$, we will now prove that we get that either $w(x, y) = w(x + m, y)$ and $w(x, y + n) = w(x + m, y + n)$ or $w(x, y) = w(x, y + n)$ and $w(x + m, y) = w(x + m, y + n)$.

For brevity, we denote $PV(B_{1,n-1}(x, y + 1)) = A$, $PV(B_{1,n-1}(x + m, y + 1)) = B$; $w(x, y) = a$, $w(x + m, y) = b$, $w(x, y + n) = c$ and $w(x + m, y + n) = d$:

c		d
A		B
a		b

For a letter a , we let v_a denote the Parikh vector of a single element a . Since $a_w(m, n) = 1$, the Parikh vectors of all $m \times n$ blocks are the same. Now

$$B_{m,n}(x, y) = B_{m,n}(x + 1, y) \quad \Rightarrow \quad v_a + A = v_b + B, \quad (1)$$

$$B_{m,n}(x + 1, y) = B_{m,n}(x + 1, y + 1) \quad \Rightarrow \quad v_c + A = v_d + B, \quad (2)$$

(1) and (2) imply that $v_a + v_d = v_b + v_c$, which means that either $a = b$ and $c = d$ or $a = c$ and $d = b$, as required.

If the (m, n) -lattice is not filled with a single letter, we have at some point (x, y) from the lattice $w(x, y) \neq w(x, y + n)$ or $w(x, y) \neq w(x + m, y)$. Consider the first case (the second one is symmetric). From what we proved above, in the first case we have $w(x, y) = w(x + m, y)$ and $w(x, y + n) = w(x + m, y + n)$ (the second option is not possible due to the conditions of the case). By a similar reasoning as in the previous paragraph applied to the point $(x + m, y)$, we get that $w(x + m, y) = w(x + 2m, y)$ and $w(x + m, y + n) = w(x + 2m, y + n)$, and extending this line of reasoning we obtain

$w(x, y) = w(x + im, y)$ and $w(x, y + n) = w(x + im, y + n)$ for any integer i . Now applying the same considerations to the point $(x + im, y + n)$, we get that either $w(x + im, y + n) = w(x + (i + 1)m, y + n)$ and $w(x + im, y + 2n) = w(x + (i + 1)m, y + 2n)$ or $w(x + im, y + n) = w(x + im, y + 2n)$ and $w(x + (i + 1)m, y + n) = w(x + (i + 1)m, y + 2n)$. Since we proved that $w(x + im, y + n) = w(x + (i + 1)m, y + n)$, in any case we have $w(x + im, y + 2n) = w(x + (i + 1)m, y + 2n)$. Extending this line of reasoning, we get $(m, 0)$ -periodicity in each of the horizontal lines of the lattice and hence in the whole lattice. In the second case in a symmetric way we get $(0, n)$ -periodicity in the lattice. Theorem is proved.

The following proposition shows that, contrary to one-dimensional case (Lemma 3), there exist aperiodic two-dimensional words that have abelian complexity 1 for some values m and n .

Proposition 6. *There exists an aperiodic two-dimensional word w with abelian complexity $a_w(m, n) = 1$ for infinitely many pairs (m, n) .*

Proof. We construct such a word of 2×2 blocks each of them being either $\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$ or $\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$, in a way that in $(4, 2)$ -lattices rooted in $(x, y) \in \{0, 1\}^2$ the word is $(4, 0)$ -periodic, and in $(4, 2)$ -lattices rooted in $(x, y) \in \{2, 3\} \times \{0, 1\}$ it is $(0, 2)$ -periodic. Now we fill in the blocks $\{0, 1\} \times \{2i, 2i + 1\}$, $i \in \mathbb{Z}$ choosing one of the two blocks for any i in arbitrary way, and continue by $(4, 0)$ -periodicity to fill into the blocks $\{4j, 4j + 1\} \times \{2i, 2i + 1\}$, $i, j \in \mathbb{Z}$. In the similar way we fill in the remaining blocks: we can fill in the blocks $\{2j + 2, 2j + 3\} \times \{0, 1\}$, $j \in \mathbb{Z}$ choosing one of the two blocks for any j in arbitrary way, and continue by $(0, 2)$ -periodicity to fill into the blocks $\{4j + 2, 4j + 3\} \times \{2i, 2i + 1\}$, $i, j \in \mathbb{Z}$. Here is an fragment of such a word, with letters from the $(4, 2)$ -lattices rooted in $(x, y) \in \{0, 1\}^2$ marked by bold:

0	1	0	1	0	1	1	0	0	1	1	0
1	0	1	0	1	0	0	1	1	0	0	1
1	0	0	1	1	0	1	0	1	0	1	0
0	1	1	0	0	1	0	1	0	1	0	1
0	1	0	1	0	1	1	0	0	1	1	0
1	0	1	0	1	0	0	1	1	0	0	1
0	1	0	1	0	1	1	0	0	1	1	0
1	0	1	0	1	0	0	1	1	0	0	1

It is not hard to see that any 4×2 rectangle has four occurrences of 0 and four occurrences of 1. Clearly, this holds true for blocks with even coordinates, since it contains two blocks of the form $\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$ or $\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$. Its shifts by $(0, 1)$, $(1, 0)$ or $(1, 1)$ have the same Parikh vector due to $(4, 0)$ or $(0, 2)$ -periodicity in corresponding lattices and the block structure.

We remark that this word has abelian complexity 1 for infinitely many values (m, n) , for example for values of the form $(4m', 2n')$, where m' and n' are any integers. It is not hard to show that its abelian complexity is at most 3. One can show that considering position of an (m, n) -block relatively to initial block and consider its shifts by one of the 8 vectors in $\{0, 1, 2, 3\} \times \{0, 1\}$. Proposition is proved.

3.2 Abelian complexity bounded by 2

We start this section with an obvious example:

Example. There exists an aperiodic two-dimensional word w with abelian complexity $a_w(m, n) = 2$ for all pairs (m, n) . Just take $w(0, 0) = 1$ and $w(i, j) = 0$ if $(i, j) \neq (0, 0)$.

Clearly, the word from the example is not recurrent. The aim of this subsection is to prove that for recurrent words the abelian complexity cannot be bounded by 2:

Theorem 7. *Let w be a two-dimensional aperiodic recurrent word. Then there exist integers m and n such that $a_w(m, n) \geq 3$.*

We first prove a few simple technical lemmas.

Lemma 8. *Let w be a one-dimensional binary word with uniform frequency of 1 equal to α , $0 < \alpha < 1$. Then for each N there exists $n \geq N$ such that $a_w(n) \geq 2$.*

Proof. If α is irrational, clearly, we cannot have abelian complexity 1 for any n , since otherwise the word would have rational uniform frequency given by $\frac{|v|_1}{n}$ for a factor v of w of length n . If α is rational, $\alpha = \frac{p}{q}$, then take n to be any prime number greater than q and N . Then the uniform frequency of 1 is $\frac{|v|_1}{n}$ for a factor v of w of length n , which cannot be equal to $\frac{p}{q}$, since n is a prime greater than q and $\alpha \neq 0, 1$.

Lemma 9. *Let w and w' be one-dimensional binary words with uniform frequencies of 1 equal to α_0 and α_1 , respectively. If $\alpha_0 \neq \alpha_1$, then there exists N such that $PV_w(n) \cap PV_{w'}(n) = \emptyset$ for each $n \geq N$.*

Proof. Follows from the definition of uniform frequency.

Proof of Theorem 7. Assume that for each m, n we have $a_w(m, n) \leq 2$. First note that taking $m = n = 1$ we get that the word is binary; we let 0 and 1 denote the letters. Taking $n = 1$, we get that in horizontal lines we have abelian complexity at most 2 for each m . This means that in each horizontal line there exists a uniform frequency of 1. Since abelian complexity is at most 2 for each n , the uniform frequency can take at most two distinct values.

First consider the case of two values. Clearly, these two values cannot be achieved in one horizontal line. So, we have one uniform frequency in some lines and another uniform frequency in other lines. Lemma 9 implies that for n big enough we must have only one Parikh vector in each line. Let α be the uniform frequency in one of the lines. If $\alpha \neq 0, 1$, then, due to Lemma 8, for some n , which can be chosen as big as needed, we must have at

least two distinct Parikh vectors, which gives at least three in total together with Parikh vector of another uniform frequency – a contradiction. Therefore, the only possible case of two distinct uniform frequencies is given by 0 and 1. It is not hard to see that lines with uniform frequency 1 can only be all-ones lines and lines with uniform frequency 0 can only be all-zero lines. Indeed, otherwise we necessarily have abelian complexity at least 3 for $n = 2$: we have factors 00 and 11 in the lines with uniform frequency 0 and 1, respectively, and we must have factor 01 or 10 if some line contains both 0 and 1. And in the case when lines with uniform frequency 1 are all-ones lines and lines with uniform frequency 0 are all-zero lines we have a periodicity vector $(1, 0)$.

Therefore, in all horizontal lines we have the same uniform frequencies of letters. Symmetrically, we also have the same uniform frequencies of letters in all vertical lines. Taking into account our assumption about abelian complexity at most 2, we have balance in $1 \times m$ and $n \times 1$ rectangles.

Now for any (m, n) consider an $m \times n$ -rectangle and a vertical stripe of width m containing it. Consider such a rectangle and its vertical shift by one. These two rectangles have a $(m - 1) \times n$ -rectangle in common (plus a $m \times 1$ -rectangle each). Since we have balance in $m \times 1$ -rectangles and abelian complexity 2 for $m \times n$ rectangles, we get balance in $m \times n$ -rectangles in each vertical stripe of width m . Symmetrically, we get balance in each horizontal stripe of height n . Now from abelian complexity 2 in $m \times n$ -rectangles we get balance in $m \times n$ -rectangles.

Now we use the characterization of balanced two-dimensional words by V. Berthé and R. Tijdeman [2]. In the paper, Corollary from the Introduction states the following:

Corollary 10 ([2]). *Two-dimensional 1-balanced words have uniform frequencies of letters, and moreover they are fully periodic unless the uniform frequency of 1-s equals 0 or 1.*

On the other hand, it is not hard to describe balanced words of uniform frequency 0 (or, symmetrically, 1). Clearly, each rectangular block must have at most one occurrence of 1 (otherwise we have unbalanceness, since in a word with uniform frequency 0 we must have an all-0 block of any size). So, the whole two-dimensional word has at most one occurrence of 1. Therefore, this is either all-0 word, or a word containing only one occurrence of 1. The first one is fully periodic, the second one is not recurrent. The theorem is proved.

In Proposition 6 we saw that there exist aperiodic words with abelian complexity at most 3 and equal to 1 for some pairs (m, n) , but the abelian complexity 1 still gives some periodicity-like structure (periodicity in lattices). The following proposition shows that we can in fact have abelian complexity at most 3 without this periodicity-like structure:

Proposition 11. *There exist recurrent aperiodic two-dimensional words that have abelian complexity 2 or 3 and are aperiodic in any lattice.*

Proof. To build such an example, we start with any binary word $x \in \{0, 1\}^{\mathbb{Z}}$. We build a

two-dimensional word w by 2×2 blocks in the following way:

$$\begin{matrix} w_{2i,2j+1} & w_{2i+1,2j+1} \\ w_{2i,2j} & w_{2i+1,2j} \end{matrix} = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \text{if } x_i = x_j, \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & \text{if } x_i \neq x_j. \end{cases}$$

For the proof we define a “domino pair” to be a pair of adjacent elements of the form $w_{2k,l}w_{2k+1,l+1}$ or $w_{k,2l+1}w_{k,2l}$. Note that each “domino pair” contains exactly one occurrence of 0 and one occurrence of 1 by construction.

To prove that abelian complexity $a_w(m, n)$ of this word is at most 3, we consider three cases:

Case 1: m and n are both odd. In this case the Parikh vector of an $(m \times n)$ -block can be divided into $\frac{mn-1}{2}$ domino pairs plus one element which can be either 0 or 1 (See Fig. 1), hence in this case $a_w(m, n) = 2$ and

$$PV(m, n) = \left\{ \left(\frac{mn-1}{2}, \frac{mn+1}{2} \right), \left(\frac{mn+1}{2}, \frac{mn-1}{2} \right) \right\}.$$

Case 2: one of the numbers m and n is odd, and the other one is even. Without loss of generality suppose that n is odd and m is even. Here we have two subcases: either $(m \times n)$ -block is situated in a point with an even second coordinate, or with an odd one. In the first subcase the $(m \times n)$ -block contains $\frac{mn}{2}$ domino pairs and hence the Parikh vector is $(\frac{mn}{2}, \frac{mn}{2})$ (See Fig. 1, Case 2a). In the other subcase the $(m \times n)$ -block contains $\frac{mn}{2} - 1$ domino pairs plus two points (See Fig. 1, Case 2b). These two points can be either both 0, or both 1, or one of them 0 and the other one 1, which gives three possible Parikh vectors, one of which coinciding with the previous subcase:

$$PV(m, n) = \left\{ \left(\frac{mn}{2}, \frac{mn}{2} \right), \left(\frac{mn}{2} - 1, \frac{mn}{2} + 1 \right), \left(\frac{mn}{2} + 1, \frac{mn}{2} - 1 \right) \right\}$$

and $a_w(m, n) = 3$.

Case 3: both m and n are even. Here we have two subcases: either $(m \times n)$ -block is situated in a point with one or both even coordinates, or both coordinates are odd. In the first subcase the $(m \times n)$ -block contains $\frac{mn}{2}$ domino pairs and hence the Parikh vector is $(\frac{mn}{2}, \frac{mn}{2})$ (See Fig. 1, Case 3a). In the other subcase the $(m \times n)$ -block contains $\frac{mn}{2} - 2$ domino pairs plus four points of the form $(2k+1, 2l+1)$, $(2k+m, 2l+1)$, $(2k+1, 2l+n)$ and $(2k+m, 2l+n)$ (See Fig. 1, Case 3b). By construction, if $x_s = x_t$, then $w_{2s,i} = w_{2t,i} = 1 - w_{2s+1,i} = 1 - w_{2t+1,i}$ for each i ; and if $x_s \neq x_t$, then $w_{2s,i} = 1 - w_{2t,i} = 1 - w_{2s+1,i} = w_{2t+1,i}$ for each i . Applying this observation to $(s, t) \in \{(k, l), (k + \frac{m}{2}, l), (k, l + \frac{n}{2}), (k + \frac{m}{2}, l + \frac{n}{2})\}$ (for example, using a straightforward case study), we get three Parikh vectors:

$$PV(m, n) = \left\{ \left(\frac{mn}{2}, \frac{mn}{2} \right), \left(\frac{mn}{2} - 2, \frac{mn}{2} + 2 \right), \left(\frac{mn}{2} + 2, \frac{mn}{2} - 2 \right) \right\}$$

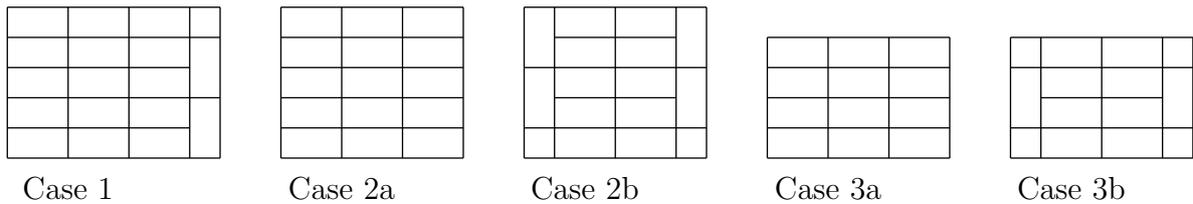


Figure 1: The cases from the proof of Proposition 11 (up to symmetry).

and $a_w(m, n) = 3$.

Summing up, the abelian complexity of this word is given by:

$$a_w(m, n) = \begin{cases} 2, & \text{if } m \text{ and } n \text{ are odd,} \\ 3, & \text{otherwise.} \end{cases}$$

It is not hard to see that if x is recurrent and aperiodic, then w is recurrent and aperiodic. The proposition is proved.

Example: Such two-dimensional word can be constructed, for example, using the Thue-Morse word t , which is defined as the fixed point of the morphism $\tau : 0 \rightarrow 01, 1 \rightarrow 10$:

$$t = 0110100110010110 \dots$$

As a word x from the above construction, we can take any word from the shift orbit closure of t , for example, $t^R t$. A fragment of the corresponding two-dimensional word of abelian complexity at most 3 is

$$\begin{array}{cccccccc} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}$$

Remark: For some values of (m, n) the balance constant is 4 for abelian complexity 3, which is not possible in one-dimensional case: in one dimension, words of abelian complexity 3 are exactly 2-balanced words (in the binary case).

3.3 Some observations on abelian complexity of one-dimensional words

By Lemma 3 we know that the abelian complexity of an aperiodic one-dimensional word is at least 2 for each n . Theorem 4 says that aperiodic words achieving the minimal abelian complexity 2 for each n are exactly Sturmian words. Can an aperiodic word have abelian

complexity greater than 2 for only one (or finitely many) values of n ? The following lemma basically says in an aperiodic non-Sturmian case the abelian complexity is not minimal for infinitely many n :

Lemma 12. *Let w be a binary aperiodic one-dimensional word which is not balanced. Then we have $a_w(n) \geq 3$ for infinitely many n .*

Proof. Since w is not balanced, there exists a length k for which at least three abelian classes exist. Since the word is aperiodic, each of its suffixes is aperiodic. Hence in its suffix for each length there are at least two abelian classes of factors, which means that for each length factors from at least two abelian classes occur infinitely many times in w . Without loss of generality assume that for length k two abelian classes of factors occurring infinitely many times contain m and $m + 1$ occurrences of a , and there also exists a factor u containing $m - 1$ occurrences of a (possibly occurring only finitely many times). We let s denote some position of an occurrence of u . Consider an occurrence of a factor v from the class with $m + 1$ occurrences of a . We let t denote the position of this occurrence, and we remark that t can be chosen bigger than any given constant. Now consider two factors $U = w_s \cdots w_{t-1}$ and $V = w_{s+k} \cdots w_{t+k-1}$ of the same length $t - s$, which can be bigger than any given constant by the remark above.

$$\begin{aligned} |U|_a - |V|_a &= (|w_s \cdots w_{s+k-1}|_a + |w_{s+k} \cdots w_{t-1}|_a) - (|w_{s+k} \cdots w_{t-1}|_a + |w_t \cdots w_{t+k-1}|_a) = \\ &= |w_s \cdots w_{s+k-1}|_a - |w_t \cdots w_{t+k-1}|_a = |u|_a - |v|_a = (m - 1) - (m + 1) = -2. \end{aligned}$$

Therefore, we have at least three abelian classes of words of length $(t - s)$. Since t can be chosen bigger than any number, we have infinitely many lengths with 3 abelian classes.

The lemma is proved.

We remark that for biinfinite words Lemma 12 does not hold, for example, the following word:

$$\cdots 0101010101011001101010101010101 \cdots$$

has three abelian classes of length 2 and two abelian classes for length greater than 2. By the definition of periodicity for biinfinite words, this word is aperiodic – although obviously it has some periodicity structure similar to ultimate periodicity (prefixes and suffixes are periodic). However, for aperiodic biinfinite words recurrence is enough to guarantee that it is impossible to have three abelian classes for finitely many lengths:

Lemma 13. *Let w be a binary aperiodic recurrent biinfinite one-dimensional word which is not balanced. Then we have $a_w(n) \geq 3$ for infinitely many n .*

Proof. Suppose the converse, and let n denote the biggest length for which we have three or more abelian classes. Consider two factors u, v of w of length n with $|u|_1 - |v|_1 \geq 2$. Since w is recurrent, u and v occur infinitely many times in w . In particular, this means that for each $N \in \mathbb{N}$, there exist an occurrence of u at position i and an occurrence of v at position j , such that $|j - i| > N$. Without loss of generality suppose that $j > i$. Then $|w_i \cdots w_{j-1}|_1 - |w_{i+n} \cdots w_{j+n-1}|_1 = |w_i \cdots w_{i+n-1}|_1 - |w_j \cdots w_{j+n-1}|_1 = |u|_1 - |v|_1 \geq 2$. Therefore, for length $j - i$ we also have at least three abelian factors. A contradiction.

3.4 Abelian complexity eventually bounded by 2

A fact similar to Lemma 13 is true also for two-dimensional words and even in a slightly stronger form (no aperiodicity is needed), although the proof is more complicated:

Theorem 14. *Let w be a binary two-dimensional recurrent word whose abelian complexity is not bounded by 2, i.e. there exists a pair (m, n) of integers with $a_w(m, n) > 2$. Then there exists infinitely many pairs (m, n) for which the inequality holds.*

The proof is split into several lemmas.

Lemma 15. *Let w be a binary two-dimensional word such that its abelian complexity is eventually bounded by 2, i.e. there exist $M, N \in \mathbb{N}$ such that for each (m, n) with $n > N$ or $m > M$ we have $a_w(m, n) \leq 2$. Then w has uniform frequencies of letters. Moreover, the balance condition is satisfied for $m \times n$ -rectangles with $n > N$ or $m > M$.*

Proof. The proof is similar to the first part of the proof of Theorem 7. Taking $n = 1$, we get that in horizontal lines we have abelian complexity at most 2 for each $m > M$. This means that in each horizontal line there exists a uniform frequency of 1. Since abelian complexity is at most 2 for each $m > M$, the uniform frequency can take at most two distinct values.

First consider the case of two values. Clearly, these two values cannot be achieved in one horizontal line. So, we have one uniform frequency in some lines and another uniform frequency in other lines. Lemma 9 implies that for m big enough we must have only one Parikh vector in each line. Let α be the uniform frequency in one of the lines. If $\alpha \neq 0, 1$, then, due to Lemma 8, for some m , which can be chosen as big as needed, we must have at least two distinct Parikh vectors, which gives at least three in total together with Parikh vector of another uniform frequency – a contradiction. Therefore, the only possible case of two distinct uniform frequencies is given by 0 and 1. It is not hard to see that lines with uniform frequency 1 can only be all-ones lines and lines with uniform frequency 0 can only be all-zero lines. Indeed, otherwise we necessarily have abelian complexity at least 3 for each $m > 2$: we have factors 0^m and 1^m in the lines with uniform frequency 0 and 1, respectively, and we must have factor 01^{m-1} or 10^{m-1} if some line contains both 0 and 1. And in the case when lines with uniform frequency 1 are all-ones lines and lines with uniform frequency 0 are all-zero lines we have a periodicity vector $(1, 0)$.

Therefore, in all horizontal lines we have the same uniform frequencies of letters and the balance condition for $m \times 1$ -rectangles with $m \geq M$. Symmetrically we prove that the uniform frequencies are the same in vertical lines, and get balance condition in vertical rectangles. The existence of uniform frequency and balance follows. The balance in $m \times n$ rectangles is proved in exactly the same way as in the Theorem 7. Lemma is proved.

Lemma 16. *Let w be a two-dimensional word satisfying the conditions of Lemma 15. Then the uniform frequencies of letters in w are rational.*

To prove that, we will make use of properties of Sturmian words and in particular their definition via rotations. Before giving the proof, we provide some facts about Sturmian

words, making use of their definition via rotations. In [19], Morse and Hedlund showed that each Sturmian word may be realized geometrically by an irrational rotation on the circle. More precisely, every Sturmian word x is obtained by coding the symbolic orbit of a point $\rho(x)$ on the circle (of circumference one) under a rotation by an irrational angle α where the circle is partitioned into two complementary intervals, one of length α (labeled 1) and the other of length $1 - \alpha$ (labeled 0). And conversely each such coding gives rise to a Sturmian word. The irrational α is called the *slope* and the point $\rho(x)$ is called the *intercept* of the Sturmian word x . More precisely, a Sturmian word $s_{\alpha,\rho}$ is defined as

$$s_{\alpha,\rho}(i) = \begin{cases} 1, & \text{if } \{\rho + i\alpha\} \in I_1, \\ 0, & \text{if } \{\rho + i\alpha\} \in I_0, \end{cases}$$

where $I_1 = (1 - \alpha, 1]$ and $I_0 = (0, 1 - \alpha]$ (or $I_1 = [1 - \alpha, 1)$ and $I_0 = [0, 1 - \alpha)$).

It is well known (e.g. [17], see also a short survey in [11]) that each of the $m + 1$ factors of a Sturmian word of length m corresponds to a half-open interval on the circle. More precisely, if we arrange the $m + 2$ points $0, 1, \{-\alpha\}, \{-2\alpha\}, \dots, \{-m\alpha\}$ in increasing order, we determine a partition of the circle into $m + 1$ half-open subintervals $L_0(m), L_1(m), \dots, L_m(m)$. Each of these subintervals is in bijection with a factor of length m of a Sturmian word of slope α . Moreover, the factors associated with these intervals are lexicographically ordered.

Recall that a factor of length m of a Sturmian word s_α has a Parikh vector equal either to $(\lceil m\alpha \rceil, m - \lceil m\alpha \rceil)$ (in which case it is called *light*) or to $(\lfloor m\alpha \rfloor, m - \lfloor m\alpha \rfloor)$ (in which case it is called *heavy*). We let $I(\alpha, \beta)$, $\alpha, \beta \in [0, 1)$, $\alpha < \beta$, denote the interval $[\alpha, \beta)$ if $0 \in I_1$ and for $(\alpha, \beta]$ if $0 \notin I_1$. The following proposition relates the intervals $L_i(m)$ to the Parikh vectors of the associated factors:

Proposition 17 (Proposition 3.3. in [11]; see also Theorem 19 in [23]). *Let s be a Sturmian word of angle α and m be a positive integer. Let t_i be the factor of length m associated with the interval $L_i(m)$. Then t_i is heavy if $L_i(m) \subset I(\{-m\alpha\}, 1)$, while it is light if $L_i(m) \subset I(0, \{-m\alpha\})$.*

Proof of Lemma 16. Suppose the uniform frequencies are irrational. We let α denote the uniform frequency of 1. Due to balance in $m \times 1$ blocks (given by Lemma 15), we have uniform frequency α in all horizontal lines. Hence all horizontal lines, as well as their one-way infinite suffixes, are aperiodic. By Lemma 12, in each line we have a balanced word with the same irrational uniform frequencies of letters, hence all the words in horizontal lines are Sturmian with the same slope.

We will now prove that in this case, for two consecutive horizontal lines (corresponding to $y = t$ and $y = t + 1$), there exists $m > M$ (where M is as in Lemma 15) and positions i and j such that $||B_{m,2}(i, t)|_0 - |B_{m,2}(j, t)|_0| = 2$ (equivalently, for any two Sturmian words of the same slope, there exist infinitely many m such that they both have at some position factors are light, and in another position factors heavy). This means that w is unbalanced in rectangle $m \times 2$ with $m > M$, which gives the unbalance contradicting Lemma 15. First note that by Kronecker's theorem (see, e.g., [13]), the points $\{-m\alpha\}$

are dense in the circle for α irrational. We now have two Sturmian words with the same slope α and different intercepts x and x' : $s_{\alpha,x}$ and $s_{\alpha,x'}$. Since $\{-m\alpha\}$ are dense in the circle, we can choose m such that the distance between x and x' on the circle (i.e., $\min(|x - x'|, 1 - |x - x'|)$) is smaller than $\min(\{-m\alpha\}, 1 - \{-m\alpha\})$. Now, again due to the fact that the points $\{-m\alpha\}$ are dense in the circle, there exists a position i such that both $x + i\alpha \in I(\{-m\alpha\}, 1)$ and $x' + i\alpha \in I(\{-m\alpha\}, 1)$. Symmetrically, there exists a position j such that both $x + j\alpha \in I(0, \{-m\alpha\})$ and $x' + j\alpha \in I(0, \{-m\alpha\})$.

Lemma is proved.

Lemma 18. *Let w be a two-dimensional word satisfying the conditions of Lemma 15, i.e. there exist $M, N \in \mathbb{N}$ such that for each (m, n) with $n > N$ or $m > M$ we have $a_w(m, n) \leq 2$, and let λ be the rational uniform frequency $\frac{p}{q}$ of the letter 1. Let (m_1, n_1) and (m_2, n_2) be two pairs of integers satisfying $m_i > M$ or $n_i > N$ for $i = 1, 2$, with $m_1 n_1$ and $m_2 n_2$ divisible by q . Then one of the three possibilities holds for pairs (m_1, n_1) and (m_2, n_2) simultaneously:*

1. $PV(m_i, n_i) = \{(m_i n_i \frac{p}{q}, m_i n_i (1 - \frac{p}{q})), (m_i n_i \frac{p}{q} + 1, m_i n_i (1 - \frac{p}{q}) - 1)\}$,
2. $PV(m_i, n_i) = \{(m_i n_i \frac{p}{q}, m_i n_i (1 - \frac{p}{q})), (m_i n_i \frac{p}{q} - 1, m_i n_i (1 - \frac{p}{q}) + 1)\}$,
3. $PV(m_i, n_i) = \{(m_i n_i \frac{p}{q}, m_i n_i (1 - \frac{p}{q}))\}$.

In particular, in the case $m_1 n_1 = m_2 n_2$ we have $PV(m_1, n_1) = PV(m_2, n_2)$.

Proof. Due to the condition on m_i and n_i , we have balance in $m_i \times n_i$ rectangles by Lemma 15. By the condition on the uniform frequency, one of the three cases holds for each pair (m_i, n_i) independently. So, it remains to prove that they must happen simultaneously.

Suppose first that (m_1, n_1) satisfies Case 1, i.e., $PV(m_1, n_1) = \{(m_1 n_1 \frac{p}{q}, m_1 n_1 (1 - \frac{p}{q})), (m_1 n_1 \frac{p}{q} + 1, m_1 n_1 (1 - \frac{p}{q}) - 1)\}$. Consider a rectangular factor of size $(m_1 m_2, n_1 n_2)$ and its factorization with $m_1 \times n_1$ -rectangles, such that in the factorization there is a rectangle with Parikh vector $(m_1 n_1 \frac{p}{q} + 1, m_1 n_1 (1 - \frac{p}{q}) - 1)$. Clearly, the uniform frequency of 1 in this bigger rectangle is larger than λ . Now consider its factorization with $m_1 \times n_1$ - and $m_2 \times n_2$ -rectangles. Due to uniform frequency and $m_2 \times n_2$ -balance, it must contain a $m_2 \times n_2$ -rectangle with Parikh vector $(m_1 n_1 \frac{p}{q} + 1, m_1 n_1 (1 - \frac{p}{q}) - 1)$. Therefore, $PV(m_2, n_2)$ also satisfies Case 1. Case 2 is symmetric.

Now suppose that (m_1, n_1) satisfies Case 3, i.e., we have only one abelian class for $m_1 \times n_1$ -rectangle: $PV(m_1, n_1) = \{(m_1 n_1 \frac{p}{q}, m_1 n_1 (1 - \frac{p}{q}))\}$. If for $m_2 \times n_2$ -rectangles we have only one abelian class, then clearly its Parikh vector should be the same. The case of two Parikh vectors for $(m_2 \times n_2)$ is impossible, since we have just proved that it implies two Parikh vectors for $(m_1 \times n_1)$. Lemma is proved.

Proof of Theorem 14. Suppose the converse: let w be a two-dimensional word which has abelian complexity bounded by 2 for all but a finite non-empty set of pairs (m, n) . Let $S \times T$ be the largest block for which $a_w(S, T) \geq 3$ (in the sense that for all blocks with $m > S$ and $n \geq T$ or $m \geq S$ and $n > T$ we have $a_w(m, n) \leq 2$). We remark that the

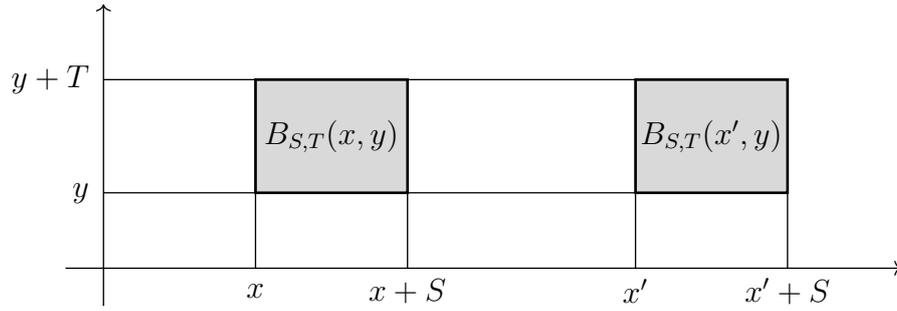


Figure 2: Illustration to the proof of Claim 1.

choice of S and T is not necessarily unique, and that for M and N defined in Lemma 15, we have $M \geq S, N \geq T$. So, there exist an integer l and occurrences of $S \times T$ -blocks with $> l$ occurrences of 1, l occurrences of 1 and $< l$ occurrences of 1. By Lemmas 15 and 16, the uniform frequency of 1 in w exists and is rational; we let λ denote this uniform frequency.

Claim 1: Let (x, y) be an occurrence of $S \times T$ -block with $> l$ occurrences of 1, i.e., $|B_{S,T}(x, y)|_1 > l$. Then for each $x' > x + M$ we have $|B_{S,T}(x', y)|_1 \geq l$. Symmetrically, the same holds for (x', y) with $x' < x - M$ and for each (x, y') with $y' > y + N$ or $y' < y - N$. And, symmetrically, one can replace $> l$ with $< l$ and obtain a similar statement.

Proof of Claim 1. Assume the converse, i.e., there exists $x' > x + M$ with

$$|B_{S,T}(x', y)|_1 < l.$$

Then

$$\begin{aligned} &||B_{x'-x,T}(x, y)|_1 - |B_{x'-x,T}(x + S, y)|_1| = \\ &||B_{S,T}(x, y)|_1 - |B_{S,T}(x', y)|_1| \geq 2 \end{aligned}$$

(see Fig. 2) So, we get unbalance in $(x' - x) \times T$ -blocks; since $x' - x > M$, we have a contradiction with Lemma 15. The other statements are symmetric. Claim 1 is proved.

Claim 2: Let (x, y) be an occurrence of $S \times T$ -block with $> l$ occurrences of 1, i.e., $|B_{S,T}(x, y)|_1 > l$. Then for each $x' > x + M + S$ we have $|B_{S,T}(x', y)|_1 \leq l$. Symmetrically, the same holds for (x', y) with $x' < x - M - S$ and for each (x, y') with $y' > y + N + T$ or $y' < y - N - T$. And, symmetrically, one can replace $> l$ with $< l$ and obtain a similar statement.

Proof of Claim 2. Assume the converse, that is, there exists $x' > x + M + S$ with $|B_{S,T}(x', y)|_1 > l$. Consider an occurrence (x_1, y_1) of an $S \times T$ -block with $< l$ occurrences of 1. Due to Claim 1 applied to (x_1, y_1) (a symmetric statement for $< l$),

$$|B_{S,T}(x_1 + x' - x, y_1)|_1 \leq l.$$

Then either

$$||B_{x'-x+S,T}(x, y)|_1 - |B_{x'-x+S,T}(x_1, y_1)|_1| \geq 2,$$

or

$$||B_{x'-x-S,T}(x + S, y)|_1 - |B_{x'-x-S,T}(x_1 + S, y_1)|_1| \geq 2.$$

So, we get unbalance in $(x' - x + S) \times T$ -blocks, or in $(x' - x - S) \times T$ -blocks; since $x' - x - S > M$, we have a contradiction with Lemma 15. The other statements are symmetric. Claim 2 is proved.

Claims 1 and 2 immediately imply

Corollary: Let l be as above, (x, y) be an occurrence of $S \times T$ -block with $> l$ or $< l$ occurrences of 1, i.e., $|B_{S,T}(x, y)|_1 \neq l$. Then for each $x' > x + M + S$ we have $|B_{S,T}(x', y)|_1 = l$. Symmetrically, the same holds for (x', y) with $x' < x - M - S$ and for each (x, y') with $y' > y + N + T$ or $y' < y - N - T$.

Claim 3: For the (rational) uniform frequency $\lambda = \frac{p}{q}$ of 1 in the word, we have $\frac{p}{q} = \frac{l}{ST}$.

Proof of Claim 3. Since the uniform frequency is $\frac{p}{q}$, due to Lemma 18, we have in the blocks $iS \times T$ for iST divisible by q the uniform frequency of 1 equal to either $\frac{p}{q} = \frac{l}{ST}$, or $\frac{p}{q} = \frac{l+1}{ST}$, or $\frac{p}{q} = \frac{l-1}{ST}$. The second case is impossible, since considering the $2iS \times T$ blocks, we get imbalance: we get blocks with $2l$ occurrences of 1's (in positions (x', y)), as well as blocks with $\frac{p}{q}2iST = 2(l+1)$ occurrences of 1's (due to uniform frequency and Lemma 18). The difference in the numbers of occurrences of 1's is 2; the unbalance follows. The third case is symmetric. The claim is proved.

Now we continue the proof of the theorem. By Lemma 18, we get that for any pair (m, n) of integers satisfying $m > M$ or $n > N$, with mn divisible by q , there are $m \times n$ blocks of uniform frequency $\lambda = \frac{p}{q}$, and moreover we have three possibilities for the sets of Parikh vectors:

1. $PV(m, n) = \{(mn\frac{p}{q}, mn(1 - \frac{p}{q})), (mn\frac{p}{q} + 1, mn(1 - \frac{p}{q}) - 1)\}$,
2. $PV(m, n) = \{(mn\frac{p}{q}, mn(1 - \frac{p}{q})), (mn\frac{p}{q} - 1, mn(1 - \frac{p}{q}) + 1)\}$,
3. $PV(m, n) = \{(mn\frac{p}{q}, mn(1 - \frac{p}{q}))\}$.

We will now show that the third case is impossible. Consider two blocks of the same size: let (x, y) be an occurrence of $S \times T$ -block with $> l$ or $< l$ occurrences of 1, i.e., $|B_{S,T}(x, y)|_1 \neq l$. Now we consider for $x' > x + M + S$ a block $B_{x'-x,T}(x, y)$ and its shift by $(S, 0)$: $B_{x'-x,T}(x + S, y)$. Due to the corollary above, we have $|B_{S,T}(x', y)|_1 = l$. Therefore,

$$|B_{x'-x,T}(x, y)|_1 - |B_{x'-x,T}(x + S, y)|_1 \neq 0.$$

So, we have distinct Parikh vectors for these blocks (Case 3 is therefore impossible), and hence for all (m, n) satisfying $m > M$ or $n > N$, with mn divisible by q , we have either

Case 1 or Case 2. Without loss of generality assume we have Case 1, i.e., $PV(m, n) = \{(mn\frac{p}{q}, mn(1 - \frac{p}{q})), (mn\frac{p}{q} + 1, mn(1 - \frac{p}{q}) - 1)\}$.

Now take any two occurrences of an $S \times T$ -block at positions (x, y) and (x', y') with $l + 1$ occurrences of 1. Due to recurrence and Corollary above, we can assume that $(x' - x)$ and $(y' - y)$ are divisible by q and up to symmetry $x' > x + M + S$, $y' > y + N + T$. Indeed, due to recurrence there are infinitely many occurrences of $S \times T$ blocks with $l + 1$ occurrences of 1. Due to the corollary above, there cannot be infinitely many with the same x or the same y . Therefore, we must have an infinite sequence of occurrences (x_i, y_i) with $x_{i+1} > x_i + M + S$, $y_{i+1} > y_i + N + T$ (up to symmetry). By the pigeonhole principle applied to residues $(x_i \bmod q, y_i \bmod q)$, we can choose two occurrences with the same residues; the divisibility follows. By the choice of x, x', y, y' we have:

$$|B_{S,T}(x, y)|_1 = |B_{S,T}(x', y')|_1 = l + 1; \tag{3}$$

Now we notice that (see yellow rectangles on Fig. 3):

$$|B_{x'-x,T}(x + S, y)|_1 = |B_{x'-x,T}(x, y')|_1 = B(x' - x)\frac{p}{q}; \tag{4}$$

Indeed, otherwise due to Lemma 18 for $(x' - x) \times T$ blocks we would have $|B_{x'-x,T}(x + S, y)|_1 = T(x' - x)\frac{p}{q} + 1$ and hence $|B_{x'-x,T}(x, y')|_1 = T(x' - x + S)\frac{p}{q} + 2$, contradicting Lemma 18 for $(x' - x + S) \times T$ blocks. Similarly for the other equality.

Due to Lemma 18 we have the following inequality (in fact, the inequality holds for all blocks of sizes satisfying the conditions of the lemma) for the green rectangle on Fig. 3:

$$|B_{x'-x+S,y'-y-T}(x, y + T)|_1 \geq (x' - x + S)(y' - y - T)\frac{p}{q}. \tag{5}$$

Summing up (3), (4) and (5), we obtain for big rectangle on Fig. 3:

$$|B_{x'-x+S,y'-y+T}(x, y)|_1 \geq (x' - x + S)(y' - y + T)\frac{p}{q} + 2,$$

which contradicts Lemma 18. The theorem is proved.

In the end of this section we treat the case of a non-binary alphabet. Clearly, for non-binary alphabet abelian complexity cannot be bounded by 2 (since it is at least 3 for 1×1 block). But it is not hard to show that the fact similar to Theorems 7 and 14 also holds for non-binary alphabet:

Theorem 19. *Let w be a two-dimensional aperiodic recurrent word over an alphabet of more than two letters. Then there exists infinitely many pairs (m, n) for which $a_w(m, n) \geq 3$.*

Proof. Suppose the converse, i.e., the abelian complexity of w is eventually bounded by 2, so that we can define $M, N \in \mathbb{N}$ to be such that for each (m, n) with $n > N$ or $m > M$ we have $a_w(m, n) \leq 2$. Consider $m > M$. Since w is aperiodic, there exists a

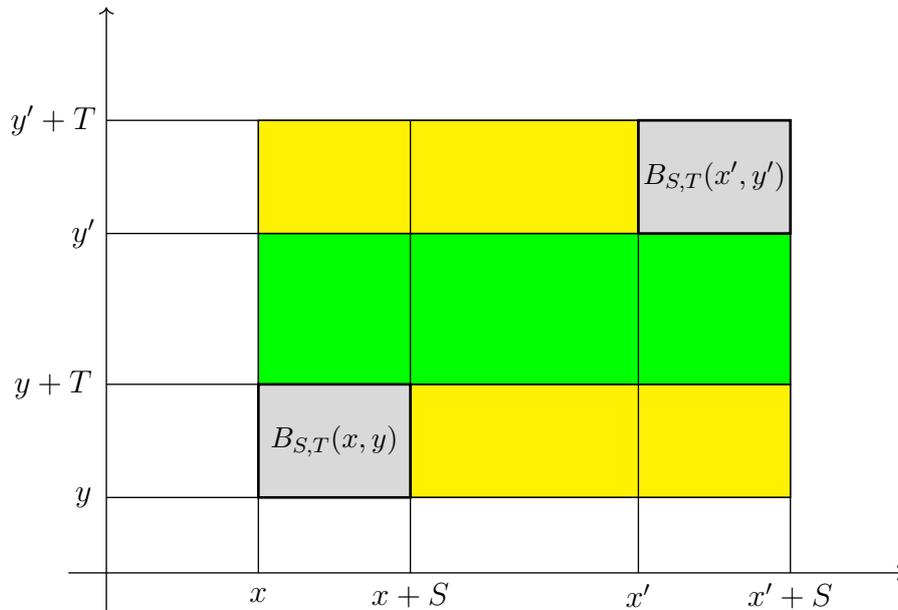


Figure 3: Illustration to the proof of Theorem 14.

letter a such that this letter is not $(m, 0)$ -periodic. Consider an occurrence (x, y) of a : $w(x, y) = a$, such that $w(x + m, y) \neq a$; we let b denote the letter at position $(x + m, y)$: $w(x + m, y) = b$. Consider Parikh vectors of blocks $B_{m,1}(x, y)$ and $B_{m,1}(x + 1, y)$. Let $PV(B_{m-1,1}(x + 1, y)) = (p_1, \dots, p_{|\Sigma|})$, hence $PV(B_{m,1}(x, y)) = (p_1, \dots, p_a + 1, \dots, p_{|\Sigma|})$; $PV(B_{m,1}(x + 1, y)) = (p_1, \dots, p_b + 1, \dots, p_{|\Sigma|})$. Since we only have two Parikh vectors for blocks of size $m \times 1$, this means that all letters distinct from a and b are $(m, 0)$ -periodic (and since we have non-binary alphabet, there is at least one such letter). Moreover, periodic letters must appear in each horizontal line (otherwise there will be a third Parikh vector). Now take any pair of relatively prime numbers (m_1, m_2) and notice that any two arithmetic progressions with differences m_1 and m_2 intersect. Hence the $(m_1, 0)$ -periodic letter is the same as m_2 -periodic letter. Now since for any integer n there exist integers p and q such that $n = pm_1 + qm_2$, we obtain that the whole word is constituted from this single letter. A contradiction.

4 Conclusions

In the paper we considered the minimal abelian complexity of two-dimensional words under the condition of recurrence. We saw that: (1) There exist aperiodic infinite words with abelian complexity 1 for infinitely many pairs (m, n) (contrary to one-dimensional case, where abelian complexity equal to 1 for some length n implies periodicity). (2) Although the minimal abelian complexity of aperiodic two-dimensional words can be equal to 2 for all pairs (n, m) , under the condition of recurrence it cannot be bounded by 2. Moreover, it must be greater than two for infinitely many pairs (n, m) . (3) There exist

uniformly recurrent words with abelian complexity bounded by 3.

An interesting open question concerns the (non-abelian) Nivat's conjecture under the condition of recurrence. It has been recently proved that if we assume that an infinite word is uniformly recurrent and for some (m, n) its complexity is bounded by mn , then it has a periodicity vector [14]. On the other hand, Julien Cassaigne characterized infinite words of complexity $mn + 1$ [9]. In particular, he showed that none of such sequences is uniformly recurrent. A challenging open problem is to find uniformly recurrent aperiodic two-dimensional words of the lowest complexity. We remark that one well-known family of two-dimensional words of low complexity include Sturmian words, i.e., words obtained by a rotation on a torus [3]: Let α, β be real numbers, with 1, α, β rationally independent, and $0 < \alpha + \beta < 1$. The two-dimensional Sturmian word s over the three-letter alphabet $\{1, 2, 3\}$ (with parameters α, β, ρ) is defined as

$$s_{m,n} = i \Leftrightarrow (m\alpha + n\beta + \rho \text{ modulo } 1) \in I_i,$$

where $I_3 = [0, \alpha), I_2 = [\alpha, \alpha + \beta), I_1 = [\alpha + \beta, 1)$ (the intervals may also be all half-open on the left). These words have no periodicity vector and they have complexity $p(m, n) = mn + m + n$. A *projection* of a two-dimensional Sturmian word defined by $v(m, n) = 1$ if and only if $s(m, n) = 1$ or 3 and $v(m, n) = 2$ if and only if $s(m, n) = 2$ gives an example of a two-dimensional uniformly recurrent word of complexity $p(m, n) = mn + n$ for sufficiently large m and n .

Acknowledgements

The author is grateful to Sergey Avgustinovich for inspiring discussions. This work has been partially supported by Russian Foundation of Basic Research (grant 18-31-00118) and by the Foundation for the Advancement of Theoretical Physics and Mathematics "BASIS".

References

- [1] S. Avgustinovich, J. Karhumäki, S. Puzynina, On abelian versions of critical factorization theorem. *RAIRO - Theor. Inf. and Applic.* 46 (2012): 3–15.
- [2] V. Berthé, R. Tijdeman, Balance properties of multi-dimensional words: *Theoret. Comput. Sci.* 273 (2002), 197–224.
- [3] V. Berthé, L. Vuillon, Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences. *Discrete Mathematics* 223 (2000): 27–53.
- [4] É. Charlier, S. Puzynina, É. Vandomme, Recurrence in Multidimensional Words. *LNCS 11417 (LATA 2019)*: 397–408.
- [5] V. Cyr, B. Kra, Nonexpansive \mathbb{Z}^2 -subdynamics and Nivat's conjecture. *Trans. Amer. Math. Soc.* 367 (2015), 6487–6537.
- [6] F. Durand, M. Rigo, Multidimensional extension of the Morse-Hedlund theorem, *European Journal of Combinatorics* 34(2), 2011.

- [7] J. Kari, M. Szabados: An algebraic geometric approach to Nivat’s conjecture. *ICALP* (2) 2015: 273–285.
- [8] M. Lothaire. *Algebraic combinatorics on words*. Cambridge University Press, 2002.
- [9] J. Cassaigne, Double Sequences with Complexity $mn + 1$. *Journal of Automata, Languages and Combinatorics* 4 (1999): 153–170.
- [10] J. Cassaigne, Subword complexity and periodicity in two or more dimensions. *Developments in Language Theory* 1999: 14–21.
- [11] G. Fici, A. Langiu, T. Lecroq, A. Lefebvre, F. Mignosi, J. Peltomäki, É. Prieur-Gaston, Abelian powers and repetitions in Sturmian words. *Theor. Comput. Sci.* 635 (2016), 16–34.
- [12] C. Epifanio, M. Koskas, F. Mignosi, On a conjecture on bidimensional words. *Theor. Comput. Sci.*, 1-3 (2003), 123–150.
- [13] G. Hardy, E. Wright, *An introduction to the theory of numbers*. Sixth edition. Oxford University Press, Oxford, 2008.
- [14] J. Kari, E. Moutot, Decidability and Periodicity of Low Complexity Tilings. [arXiv:1904.01267](https://arxiv.org/abs/1904.01267).
- [15] M. Lothaire, *Algebraic combinatorics on words*. Cambridge University Press, 2002.
- [16] B. Madill, N. Rampersad, The abelian complexity of the paperfolding word. *Discrete Mathematics* 313 (2013): 831–838.
- [17] F. Mignosi, Infinite Words with Linear Subword Complexity. *Theoret. Comput. Sci.*, 65(1989): 221–242.
- [18] M. Morse, G. Hedlund, Symbolic dynamics, *Amer. J. Math.* 60 (1938), 815–866.
- [19] M. Morse, G. Hedlund, Symbolic dynamics II: Sturmian sequences. *Amer. J. Math.* 62 (1940), 1–42.
- [20] M. Nivat, Invited talk at ICALP’97.
- [21] S. Puzynina, L. Q. Zamboni, Abelian returns in Sturmian words. *J. Comb. Theory, Ser. A* 120(2) (2013), 390–408.
- [22] G. Richomme, K. Saari, L. Q. Zamboni, Abelian complexity of minimal subshifts. *J. London Math. Society* 83 (2011), 79–95.
- [23] M. Rigo, P. Salimov, E. Vandomme, Some properties of abelian return words. *J. Integer Seq.*, 16:13.2.5, 2013.
- [24] A. Quas, L. Q. Zamboni, Periodicity and local complexity. *Theor. Comput. Sci.*, 319(1-3): 229–240, 2004.
- [25] A. Saarela, Ultimately Constant Abelian Complexity of Infinite Words. *Journal of Automata, Languages and Combinatorics* 14 (2009), 255–258.
- [26] R. Tijdeman, Periodicity and Almost-Periodicity. In: Gyri E., Katona G.O.H., Lovsz L., Fleiner T. (eds) *More Sets, Graphs and Numbers*. Bolyai Society Mathematical Studies, vol 15. (2006), Springer, Berlin, Heidelberg.